4. (Re)training the network

Configure your MobileNet

The retrain script can retrain either Inception V3 model or a MobileNet. In this exercise, we will use a MobileNet. The principal difference is that Inception V3 is optimized for accuracy, while the MobileNets are optimized to be small and efficient, at the cost of some accuracy.

Inception V3 has a first-choice accuracy of 78% on ImageNet, but is the model is 85MB, and requires many times more processing than even the largest MobileNet configuration, which achieves 70.5% accuracy, with just a 19MB download.

Pick the following configuration options:

- Input image resolution: 128,160,192, or 224px. Unsurprisingly, feeding in a higher resolution image takes more processing time, but results in better classification accuracy. We recommend 224 as an initial setting.
- The relative size of the model as a fraction of the largest MobileNet: 1.0, 0.75, 0.50, or 0.25. We recommend 0.5 as an initial setting. The smaller models run significantly faster, at a cost of accuracy.

With the recommended settings, it typically takes only a couple of minutes to retrain on a laptop. You will pass the settings inside Linux shell variables. Set those shell variables as follows:

```
IMAGE_SIZE=224
ARCHITECTURE="mobilenet_0.50_${IMAGE_SIZE}"
```

The graph below shows the first-choice-accuracies of these configurations (y-axis), vs the number of calculations required (x-axis), and the size of the model (circle area).

16 points are shown for mobilenet. For each of the 4 model sizes (circle area in the figure) there is one point for each image resolution setting. The 128px image size models are represented by the lower-left point in each set, while the 224px models are in the upper right.

Other notable architectures are also included for reference. "GoogleNet" in this figure is "Inception V1" in this table. An extended version of this figure is available in slides 84-89 here.